

Understanding Deep Representations through Random Weights

Yao Shu^{1*}, Man Zhu^{1*}, Kun He^{1*†}, John E. Hopcroft², Pan Zhou¹

¹ Huazhong University of Science and Technology, China

² Cornell University, USA

Abstract

We systematically study the deep representation of random weight CNN (convolutional neural network) using the DeCNN (deconvolutional neural network) architecture. We first fix the weights of an untrained CNN, and for each layer of its feature representation, we train a corresponding DeCNN to reconstruct the input image. As compared with the pre-trained CNN, the DeCNN trained on a random weight CNN can reconstruct images more quickly and accurately, no matter which type of random distribution for the CNN’s weights. It reveals that every layer of the random CNN can retain photographically accurate information about the image. We then let the DeCNN be untrained, i.e. the overall CNN-DeCNN architecture uses only random weights. Strikingly, we can reconstruct all position information of the image for low layer representations but the colors change. For high layer representations, we can still capture the rough contours of the image. We also change the number of feature maps and the shape of the feature maps and gain more insight on the random function of the CNN-DeCNN structure. Our work reveals that the purely random CNN-DeCNN architecture substantially contributes to the geometric and photometric invariance due to the intrinsic symmetry and invertible structure, but it discards the colorimetric information due to the random projection.

1 Introduction

Image representations for computer vision include conventional methods, such as SIFT [Lowe, 2004], HOG [Dalal and Triggs, 2005], Fisher Vectors [Perronnin and Dance, 2007] and sparse encoding [Yang *et al.*, 2010], as well as deep neural networks, particularly the Convolutional Neural Networks (CNNs). In recent years, various CNNs, including AlexNet [Krizhevsky *et al.*, 2012], VGG [Simonyan and Zisserman, 2015] and ResNet [He *et al.*, 2016a], have

shown great success in computer vision, especially in large-scale image and video recognition [Zeiler and Fergus, 2014; Sermanet *et al.*, 2014; Simonyan and Zisserman, 2014]. However, CNNs are designed empirically using hyper parameters and millions of the weight parameters are learned automatically by training, which to us is a “black box”. Understanding the image representations of the deep networks is far from satisfactory.

Up until recently, a few methods are presented to understand the deep representations of neural networks [Pan *et al.*, 2016; Zhuang *et al.*, 2015]. Inverting techniques are developed to understand the image representations by reconstructing the image [Mahendran and Vedaldi, 2015; Dosovitskiy and Brox, 2016]. [Dosovitskiy and Brox, 2016] proposes a deconvolutional approach based on deconvolutional neural network (DeCNN) to reconstruct images from feature representations learned from a pre-trained deep CNN, and found that features in higher layers preserve colors and rough contours of the images and discard information irrelevant for the classification task that the convolutional model is trained on. As there is no back propagation, their reconstruction is much quicker than the inverting method based on gradient descent [Mahendran and Vedaldi, 2015].

Besides, there is a growing interest in studying the untrained, random weight CNNs. Some researchers find that certain feature learning architectures can yield useful features for classification even with untrained random weights. And random weights perform only slightly worse than pre-trained weights on an one-layer convolutional pooling architecture [Jarrett *et al.*, 2009]. [Saxe *et al.*, 2011] finds that certain convolutional pooling architectures with random weights are inherently frequency selective and translation invariant, and argue that these properties underlie their performance. To understand the deep representations of untrained CNNs, [He *et al.*, 2016b] successfully accomplishes three deep visualization tasks (images inversion, texture synthesize and artistic style image generation) using untrained, random weight CNNs. [Mongia *et al.*, 2016] provides an initial analysis on why one-layer CNNs with random weights can successfully generate texture.

In this paper, we study the deep representations of untrained CNN using the DeCNN architecture. We randomly initialize and fix the weights of the CNN model, then for the random feature representations of each CNN layer, we train

*First three authors indicate equal contribution.

†Corresponding author, brooklet60@hust.edu.cn.

a corresponding DeCNN in order to reconstruct the image. We build the DeCNN architecture using the inverted layer sequence of CNN as in [Dosovitskiy and Brox, 2016]. Compared with the inversion on pre-trained CNN [Dosovitskiy and Brox, 2016], for every convolutional layer of AlexNet or VGG architecture, our approach can train the corresponding DeCNN more quickly and reconstruct the image with higher quality. It shows that the random features well preserve almost all geometric, photometric, and colorimetric information for the input image, and the reconstruction quality only decays a little for high convolutional layers. Also, the reconstruction quality is higher on VGG than on AlexNet for the same convolutional layer $\text{Conv}k$ ($k \in \{1, \dots, 5\}$), even though $\text{Conv}k$ for VGG is actually deeper than $\text{Conv}k$ for AlexNet. The reconstructed image is just a little bit blurry on $\text{Conv}5$ of AlexNet. We argue that the success of inverting the images is not because the pre-trained CNN has learned useful features for the classification, but mainly due to the CNN architecture itself that could contain all information of the image even after multiple layers random projections.

Then, we raise an interesting question: what if we also use an untrained, random weight DeCNN for the reconstruction? Now, the overall CNN-DeCNN is totally random. We do experiments on the VGG architecture. Surprisingly, we can reconstruct all position information of the image and even its luminance fluctuations for low layer representations ($\text{Conv}1$ to $\text{Conv}3$) but the colors change. For high layer representations ($\text{Conv}4$, $\text{Conv}5$), the reconstructed images are very blurry, but we can still capture the rough contours. Note that for $\text{Conv}5$, the image information passes the whole CNN-DeCNN with a total of $13 + 13$ convolutional layers.

We also explore more on the shape size and the number of feature maps to gain more insight on the random CNN-DeCNN architecture. The shape reduction on feature maps will result in randomness and blur on the reconstructed image due to the representation compression. While similarly to the boosting method, the increasing number of feature maps, with each feature map as an independent random feature projection, can promote the robustness on the reconstruction quality.

Our work provides more insight in understanding deep convolutional networks: what contributes to the training and what contributes to the architecture itself? We wish our work inspire more works in exploring the property of untrained, random weight deep networks.

2 Method

In this section, we first provide the CNN-DeCNN architecture in details, then present the DeCNN training method for the first task. And finally we describe various random distributions we used for either the random CNN or the random DeCNN.

2.1 Network Architecture

In what follows, when we say “feature representations” or “image representations”, we mean the feature vectors after the convolutional layer and the activation layer but before the pooling layer. A convolutional layer is usually followed by a

pooling layer, except for the last convolutional layer ($\text{Conv}5$ in VGG16 or AlexNet). For consistency, the output after the convolutional layer and the activation layer are regarded as the deep feature representation.

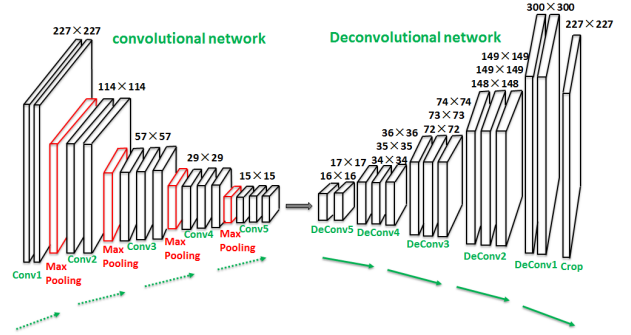


Figure 1: The CNN-DeCNN network structure of VGG16-Conv5. The convolutional network contains all convolutional layers from $\text{Conv}1$ to $\text{Conv}5$. We show input or output shape for brevity.

The overall architecture consists of a convolutional neural network (CNN) and a deconvolutional neural network (DeCNN). We select two classical CNNs, VGG16 and AlexNet, for the inversion of the feature representations due to their excellent performance on the ImageNet Large Scale Visual Recognition Competition (ILSVRC) [Deng *et al.*, 2009]. The DeCNN, also called an ‘up-convolutional’ network, is similar in structure with [Zeiler *et al.*, 2010]. For each layer of the feature representations in CNN, we build a corresponding DeCNN which combines up-sampling and convolution to do the inversion operation. We usually up-sample a feature map by factor 2: replace each value in a feature map by a 2×2 block, put the original value to the top left corner of the block and set the other three entities to be zero. Our up-sampling is the same as Dosovitskiy’s design but they only applied on pre-trained AlexNet [Dosovitskiy and Brox, 2016]. The Convolution, operation of the deconvolutional layer in the DeCNN, is the same as the convolution operation in the CNN. [Dosovitskiy and Brox, 2016] shows that the reconstructed image from the fully connected layers becomes very vague for AlexNet. As VGG16 is much deeper than AlexNet and the training for the fully connected layers takes much longer time, in this paper we will focus on the representations of the convolutional layers and explore their properties.

Figure 1 illustrates a VGG16 $\text{Conv}5$ -DeConv5 architecture, where $\text{Conv}5$ indicates the sequential layers from $\text{Conv}1$ to $\text{Conv}5$. The main idea is that the CNN and the DeCNN are symmetric and that the DeCNN is just the inverted layer sequence of the CNN. While the convolutional layer includes a convolution operation and a pooling operation, each deconvolutional layer includes an up-sampling operation and a convolution operation. Besides, each deconvolutional layer is followed by the activation layer, in which we apply the leaky ReLU nonlinearity with slope 0.2, that is, $r(x) = x$ if $x \geq 0$ and $r(x) = 0.2x$ if $x < 0$. The final Crop layer is to cut the output of DeConv1 to the same shape of the original images.

2.2 DeCNN Training

For our first task, we fix the random weights of the CNN and train the corresponding DeCNN to minimize the pixel-wise loss on the reconstructed image. Let $\Phi_i(x_i, w)$ represent the reconstruction of the DeCNN, in which x_i is the input of the i th image and w the weights of the DeCNN. We minimize the loss function such that the reconstructed image is as accurate as possible to the original image. By training the DeCNN we get the desired weights w^* that minimize the loss:

$$w^* = \arg \min_w L = \arg \min_w \sum_i (\Phi_i(x_i, w) - x_i)^2 \quad (1)$$

We initialize the DeCNN by the “MSRA” method [He *et al.*, 2015] based on a modified Caffe [Jia *et al.*, 2014] proposed in [Dosovitskiy and Brox, 2016]. We use the training data set of ImageNet [Deng *et al.*, 2009] and the Adam [Kingma and Ba, 2014] optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and the mini-batch size is set to 32. The initial learning rate is set to 0.0001 and the learning rate gradually decays by the “multistep” training. The weight decay is set to 0.0004 to avoid over-fit. As the loss has already converged after 200,000 iterations in the experiments, we set the maximum iterations as 200,000.

2.3 Random Distributions

For the random weights assigned to CNN or DeCNN, we try several Gaussian distribution with zero mean and various variance ($\delta \in \{1, 0.1, 0.015\}$) to see if they have different impact on the DeCNN’s reconstruction.

We also try different types of random distribution: Uniform, Logistic, Laplace and Gaussian to study their impact. The Uniform distribution is in $[-0.04, 0.04]$, such that the interval equals $[\mu - 3\delta, \mu + 3\delta]$ where $\mu = 0$ and $\delta = 0.015$ are parameters for the Gaussian distribution. The Logistic distribution is 0-mean and 0.015-scale and the Logistic distribution is 0-mean and 0.015-scale of decay.

Figure 2 shows the probability distributions of the random weights that we used when assigning the random weights to the CNN or DeCNN. Gaussian, Laplace and Logistic are similar in the bell curves, and their probability densities are concentrated around zero.

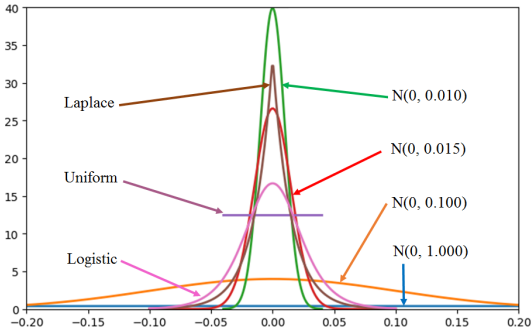


Figure 2: Probability density of the Uniform, Gaussian, Laplace and Logistic distribution of the random weights.

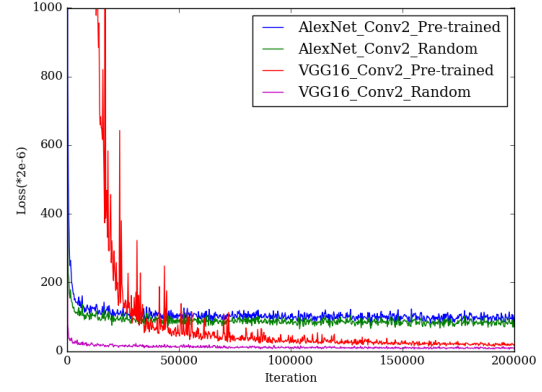


Figure 3: The training loss for Conv2-DeConv2 architecture of VGG16 and AlexNet on pre-trained or Gaussian random weight CNNs. The training converges much quicker and has slightly lower loss on the reconstruction for the random CNNs. The trend is more apparent on VGG16.

3 Experiments

We apply three types of experiments to gain insight on the deep representations of convolutional networks. We first compare the performance of Gaussian random weights of CNN and the pre-trained weights of CNN by their training loss and reconstruction quality for both VGG16-DeCNN and AlexNet-DeCNN. Then we assign random weights in different types of distribution on VGG16 to study their reconstruction quality. Finally, we assign random weights on both CNN and DeCNN of VGG16 to explore the purely random reconstruction. We also change the shape of the feature maps as well as the number of feature maps to explore their impact. All images used are from validation set of ImageNet, and some are outside ImageNet.

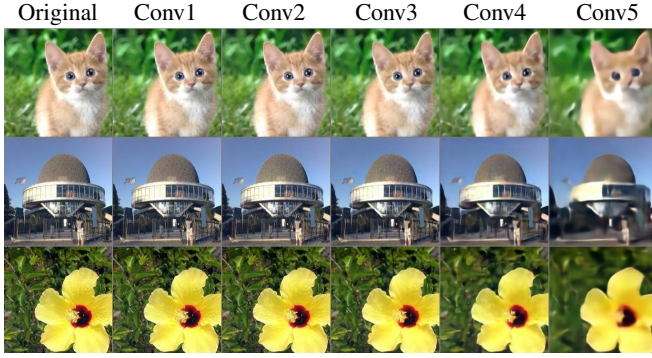
3.1 Random Weights vs. Pre-trained Weights

Let the convolutional part be the corresponding portion of VGG16 or AlexNet, depending on which layer of representation we want to reconstruct. We assign and fix random weights in $N(0, 0.015)$ Gaussian distribution to the CNN, then we initialize and train the DeCNN. By comparison, we build another CNN-DeCNN and let the weights be the pre-trained ones provided by Caffe. We also initialize and train the DeCNN using the same loss function.

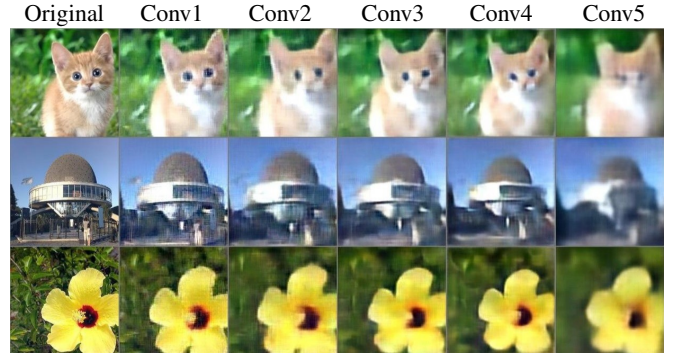
The loss curves during the training process for the Conv2-DeConv2 architecture are shown in Figure 3. The training on DeCNN converges much quicker for the random CNN and yields slightly lower loss in the end. The trend is more apparent on VGG16.

In Figure 3, we also see that the random VGG yields much more lower loss as compared with that of the random AlexNet, indicating that the VGG network structure is better than the AlexNet network structure. This may tell us that the DeCNN may provide a possible way to evaluate network structure and the hyperparameters before the training.

Figure 5 illustrates the reconstructed images on a cat image. On VGG16, the reconstructed images on either the pre-trained or random CNN are as good as the original image and



(a) Reconstruction on the rwVGG16



(b) Reconstruction on the rwAlexNet

Figure 4: Reconstruction images on VGG16 and AlexNet with random weights.

the difference is almost indistinguishable for naked eyes. On AlexNet, however, there is a considerable gap between the reconstructed image and the original image, and the reconstructed image is blurry.

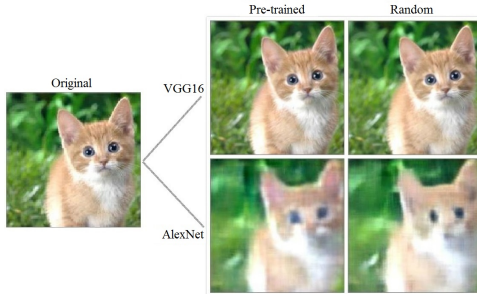


Figure 5: Illustration of reconstruction quality on the cat image for the Conv2-DeConv2 architecture. Pre-trained CNN and random CNN show similar results on the same CNN network structure. The reconstruction of DeConv on VGG16 apparently outperforms that on AlexNet.

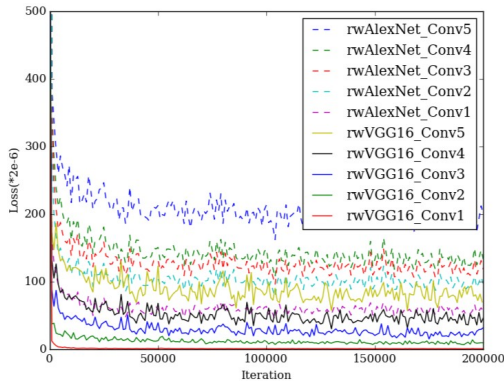


Figure 6: DeCNN training loss for rwVGG16 and rwAlexNet from Conv1 to Conv5 representations.

Figure 4 illustrates the reconstructed images from the representations of different layers on VGG16 and AlexNet, but only for random weight CNN models, denoted by rwVGG16 and rwAlexNet respectively. Here $\text{Conv}k$ represents a $\text{Conv}k$ -DeConv k architecture. We see that, on both rwVGG16 and rwAlexNet, the reconstruction quality decays for the representations of deeper layers. And the rwVGG16 network structure yields more accurate reconstruction, even

on Conv5, which involves 26 times of convolution and 4 times of max pooling operation.

Furthermore, Figure 6 shows that for the same layer, rwVGG16 architecture converges more quickly and with lower reconstruction loss than rwAlexNet. Here $\text{Conv}k$ also represents a $\text{Conv}k$ -DeConv k architecture.

Based on the above discussion, we see that random weight CNN can speed up the training process of DeCNN on both VGG16 and AlexNet. And the reconstruction quality on VGG16 is higher than the quality on AlexNet. We will focus on VGG16 and random weights in the following experiments.

3.2 Experiments on Random Distribution

In this subsection, we further explore whether different Gaussian distributions or different types of random distributions have different impact on the reconstruction quality. We did the following experiments:

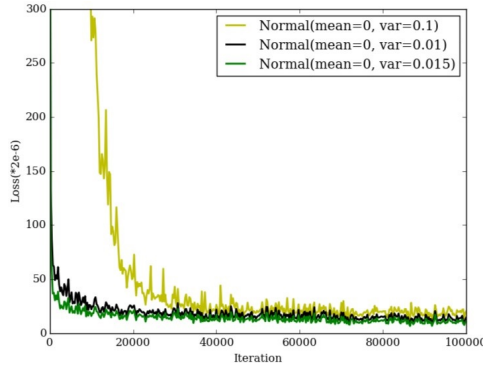
- 1) We assign different Gaussian random weights on the VGG16-CNN and train the VGG16-DeCNN.
- 2) We assign different types of random weights (Uniform, Logistic and Laplace) on the VGG16-CNN and train the VGG16-DeCNN.

Figure 7(a) shows the reconstruction loss for different Gaussian distributions. We do not show $N(0, 1)$ as it does not converge to a low value. The loss value and convergence speed are better for random distribution with small variance. But they can all converge to the same loss value.

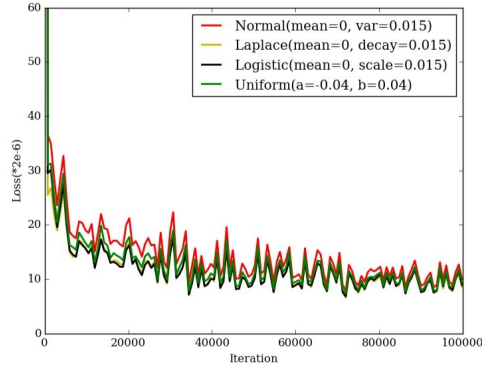
Figure 7(b) shows the reconstruction loss for different types of distribution. The loss curves nearly coincide with each other. It shows different types of random distribution work similarly for the reconstruction on rwVGG16.

Figure 8 illustrates the reconstructed images. They all reconstruct the original image very well except for $N(0, 1)$. The type of distribution has little impact on the reconstruction quality when the parameter values are picked properly.

In summary, we do not need to choose pre-trained weights or other particular weights to reconstruct images. The method with random we show is much quicker and convenient for the image reconstruction. Regarding weights in the convolutional part as an encoding method on the original image, then our network architecture can decode from the representations encoded by various methods. This may due to that the network architecture is naturally symmetric and invertible.



(a) On different Gaussian random weights



(b) On different types of random weights

Figure 7: Training loss of different types and parameters of random weights in convolutional part. Only the DeConv2 training loss for rwVGG16 is shown for brevity.

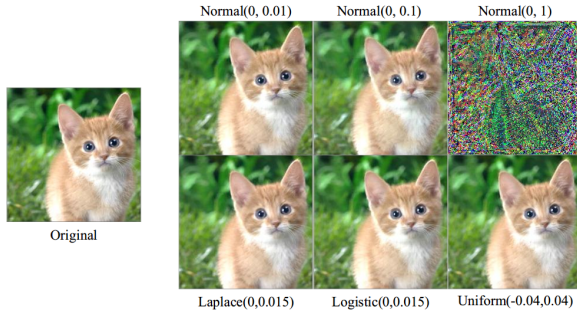


Figure 8: Image reconstructions of different types and parameters of random weights in CNN. Only the reconstructions from VGG16 Conv2-DeConv2 architecture are shown for brevity.

3.3 Random Weights for the DeCNN Network

For different types of random weights in CNN, the trained DeCNN shows excellent reconstruction quality. What if we also use untrained random weights for the DeCNN? In this subsection, we study the total random VGG16 CNN-DeCNN architecture and gain surprising insight.

We first study the reconstruction quality for different convolutional layers, as shown in Figure 9. The weights are random from $N(0, 0.1)$ and Conv k indicates a Conv k -DeConv k architecture. We see that the deeper the random representations are, the coarser the reconstructed images are. But surprisingly, even there is no train, the DeCNN can reconstruct geometric positions and contours very well. We can still perceive and guess the geometric positions of objects in images from the Conv4 representation, which is already 10 layers deep.

As the reconstruction quality decays quickly for deep layers, we argue that it may be due to the shape reduction of feature maps for higher layers. As shown in Table 1, the shape of feature maps, while going through convolutional layers, will be reduced by 1/4 except for going through the input Data layer to Conv1 layer. Due to the total randomness of the weights, the convolutional layer will project feature maps of the previous layer to a 1/4 scale shape. So the representations encoded in feature maps will be compressed and it will

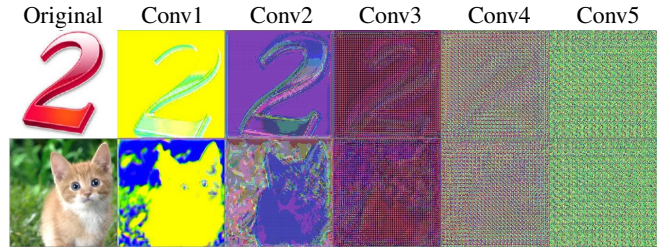


Figure 9: Reconstructed images for the total random VGG16 CNN-DeCNN, using random representations of different layers for the reconstruction. Weights are randomly generated using $N(0, 0.1)$ distribution. The deeper the CNN is, the more randomness on the reconstructed images.

be hard for a random weight DeCNN to extract these feature representations. However, the trained DeCNN can extract these representations easily and reconstruct images very well as shown in Subsection 3.1.

Layer	number of feature maps	Shape of feature maps
Data	3	227×227
Conv1	64	227×227
Conv2	128	114×114
Conv3	256	57×57
Conv4	512	29×29
Conv5	512	15×15

Table 1: The number and shape of feature maps out from each layer for VGG16 Conv5 architecture. From Conv1 to Conv5, the number of feature maps doubles and the shape of feature maps is reduced by 1/4 between two successive layers.

To get a clearer view on the impact of the shape of the feature maps, two more reconstructions of a simplified VGG16 CNN-DeCNN architecture are shown in Figure 10. Here, we simplify VGG16 CNN-DeCNN architecture by connecting the Data layer directly to Conv k followed by DeConv k and ignore other layers. Conv k in Figure 10(a) will generate the same shape and the same number of feature maps as shown in Table 1, while Conv k in Figure 10(b) will generate feature maps in the same size of the Data layer but it still generates the same number of feature maps as shown in Table 1.

In Figure 10(a), the reconstruction quality is close to that

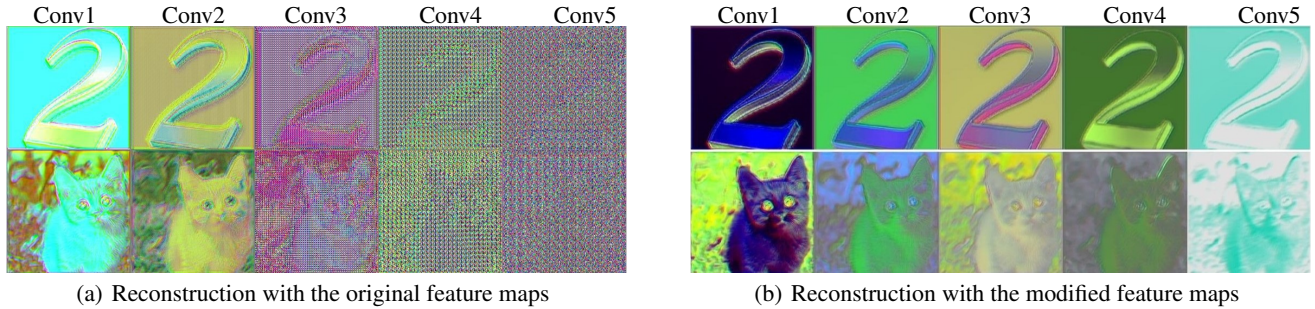


Figure 10: Reconstructed images for the simplified random VGG16 CNN-DeCNN architecture. Here $\text{Conv}k$ indicates that data layer is directly connected to the $\text{Conv}k$ followed by the corresponding $\text{DeConv}k$. Weights are from $N(0, 0.1)$ distribution. The reconstruction quality in (a) is close to the one in Figure 9, while it shows higher quality in (b).

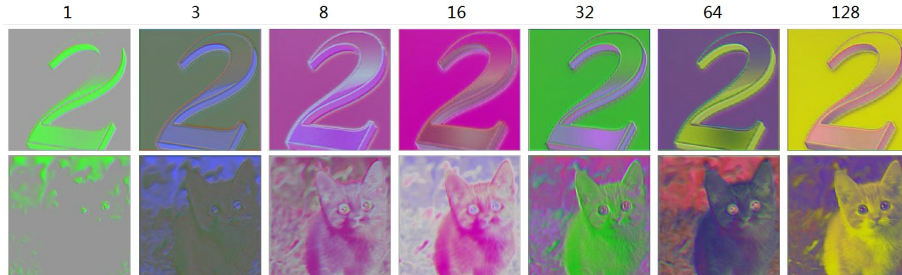


Figure 11: Reconstructed images for the simplified random VGG16 $\text{Conv}1_1$ - $\text{DeConv}1_1$ architecture. The weights are from $\text{Uniform}(-0.1, 0.1)$ distribution. The more feature maps there are, the more details are shown in the reconstruction.

in Figure 9 even though the feature representations just go through two or three layers. Surprisingly, the reconstruction quality in Figure 10(b) is the best. Even reconstructed from the representation of $\text{Conv}5$, the geometric positions and contours are clear enough for naked eyes. With the same shape of feature maps as the Data layer, the reconstructions gap from these five different convolutional layers is hard to distinguish. The shape reduction of feature maps is the key reason for the decaying of the reconstruction quality in Figure 9.

We further explore the impact for the number of feature maps using simpler architecture. We use $\text{Conv}1_1$ - $\text{DeConv}1_1$ architecture and the random weights follow the $\text{Uniform}(-0.1, 0.1)$ distribution. As shown in Figure 11, the more feature maps there are, the more details shown in the reconstruction. The increasing number of feature maps promotes the robustness of reconstruction from random DeCNN. Due to the randomness of the weights, three 227×227 vectors are randomly projected to a 227×227 space, the representations will be reformed but not be compressed. Different feature maps are independent and complementary with each other and will result in various projections, which can be merged into a much better representation in order to reconstruct the image. Similar to the boosting method, the more the feature maps there are, the more the robustness the reconstruction is.

In summary, applying random weights in the whole CNN-DeCNN architecture, we can still capture the geometric positions and contours of the image. The shape reduction of feature maps takes responsibility for the randomness on the reconstructed images for higher layer representation due to the representation compression. And random weight DeCNN can reconstruct robust images if we have enough number of

feature maps.

4 Conclusion and Discussion

In this paper we do deep visualization using deconvolutional networks to study the random representations of untrained, random weight convolutional networks.

By inverting the image representations with DeCNN, we have shown that this yields accurate reconstructions of the original image even for high-convolutional-layer representations. It shows that after the multi-layers random projection followed by convolution, pooling and nonlinear rectifier activation, the random representation can retain photographically accurate information about the image, even better than that of the pre-trained CNN. The reconstruction on VGG is with higher quality than that on AlexNet, indicating that DeCNN may provide a visualization tool to evaluate different CNN structure before the costly training.

Let the DeCNN be untrained also, for low layer representations we can surprisingly reconstruct the image with accurate geographic location but colors change, and for high layer representations we can still capture the rough contours of the image. Our work reveals that it is mainly due to the intrinsic symmetric and invertible structure of the CNN-DeCNN architecture, that with purely random weights it can gain geographic information with different degrees of geometric and photometric invariance, but it discards the colorimetric information due to the random projection and convolution.

Our work provides insight on the inner work of CNN, and through visualization to support why random weight CNNs works considerably well on image classification and why it can amazingly do texture synthesis and style transfer as shown by recent researches in the literature.

Acknowledgments

The work is supported by US Army Research Office (No. W911NF-14-1-0477), National Science Foundation of China (61472147, 61401169) and MSRA Collaborative Research (97354136).

References

- [Dalal and Triggs, 2005] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *CVPR*, volume 1, pages 886–893, 2005.
- [Deng *et al.*, 2009] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Fei-Fei Li. Imagenet: A large-scale hierarchical image database. In *CVPR*, pages 248–255, 2009.
- [Dosovitskiy and Brox, 2016] Alexey Dosovitskiy and Thomas Brox. Inverting visual representations with convolutional networks. In *CVPR*, pages 4829–4837, 2016.
- [He *et al.*, 2015] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *ICCV*, pages 1026–1034, 2015.
- [He *et al.*, 2016a] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.
- [He *et al.*, 2016b] Kun He, Yan Wang, and John Hopcroft. A powerful generative model using random weights for the deep image representation. In *NIPS*, pages 631–639, 2016.
- [Jarrett *et al.*, 2009] Kevin Jarrett, Koray Kavukcuoglu, Marc’Aurelio Ranzato, and Yann LeCun. What is the best multi-stage architecture for object recognition? In *ICCV*, pages 2146–2153, 2009.
- [Jia *et al.*, 2014] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *ACM MM*, pages 675–678, 2014.
- [Kingma and Ba, 2014] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2014.
- [Krizhevsky *et al.*, 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105, 2012.
- [Lowe, 2004] David G Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.
- [Mahendran and Vedaldi, 2015] Aravindh Mahendran and Andrea Vedaldi. Understanding deep image representations by inverting them. In *CVPR*, pages 5188–5196, 2015.
- [Mongia *et al.*, 2016] Mihir Mongia, Kundan Kumar, Akram Erraqabi, and Yoshua Bengio. On random weights for texture generation in one layer neural networks. *CoRR*, abs/1612.06070, 2016.
- [Pan *et al.*, 2016] Yingwei Pan, Yehao Li, Ting Yao, Tao Mei, Houqiang Li, and Yong Rui. Learning deep intrinsic video representation by exploring temporal coherence and graph structure. In *IJCAI*, pages 3832–3838, 2016.
- [Perronnin and Dance, 2007] Florent Perronnin and Christopher Dance. Fisher kernels on visual vocabularies for image categorization. In *CVPR*, pages 1–8, 2007.
- [Saxe *et al.*, 2011] Andrew Saxe, Pang W Koh, Zhenghao Chen, Maneesh Bhand, Bipin Suresh, and Andrew Y Ng. On random weights and unsupervised feature learning. In *ICML*, pages 1089–1096, 2011.
- [Sermanet *et al.*, 2014] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. In *ICLR*, 2014.
- [Simonyan and Zisserman, 2014] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. In *NIPS*, pages 568–576, 2014.
- [Simonyan and Zisserman, 2015] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015.
- [Yang *et al.*, 2010] Jianchao Yang, Kai Yu, and Thomas Huang. Supervised translation-invariant sparse coding. In *CVPR*, pages 3517–3524, 2010.
- [Zeiler and Fergus, 2014] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *ECCV*, pages 818–833, 2014.
- [Zeiler *et al.*, 2010] Matthew D Zeiler, Dilip Krishnan, Graham W Taylor, and Rob Fergus. Deconvolutional networks. In *CVPR*, pages 2528–2535, 2010.
- [Zhuang *et al.*, 2015] Fuzhen Zhuang, Xiaohu Cheng, Ping Luo, Sinno Jialin Pan, and Qing He. Supervised representation learning: Transfer learning with deep autoencoders. In *IJCAI*, pages 4119–4125, 2015.

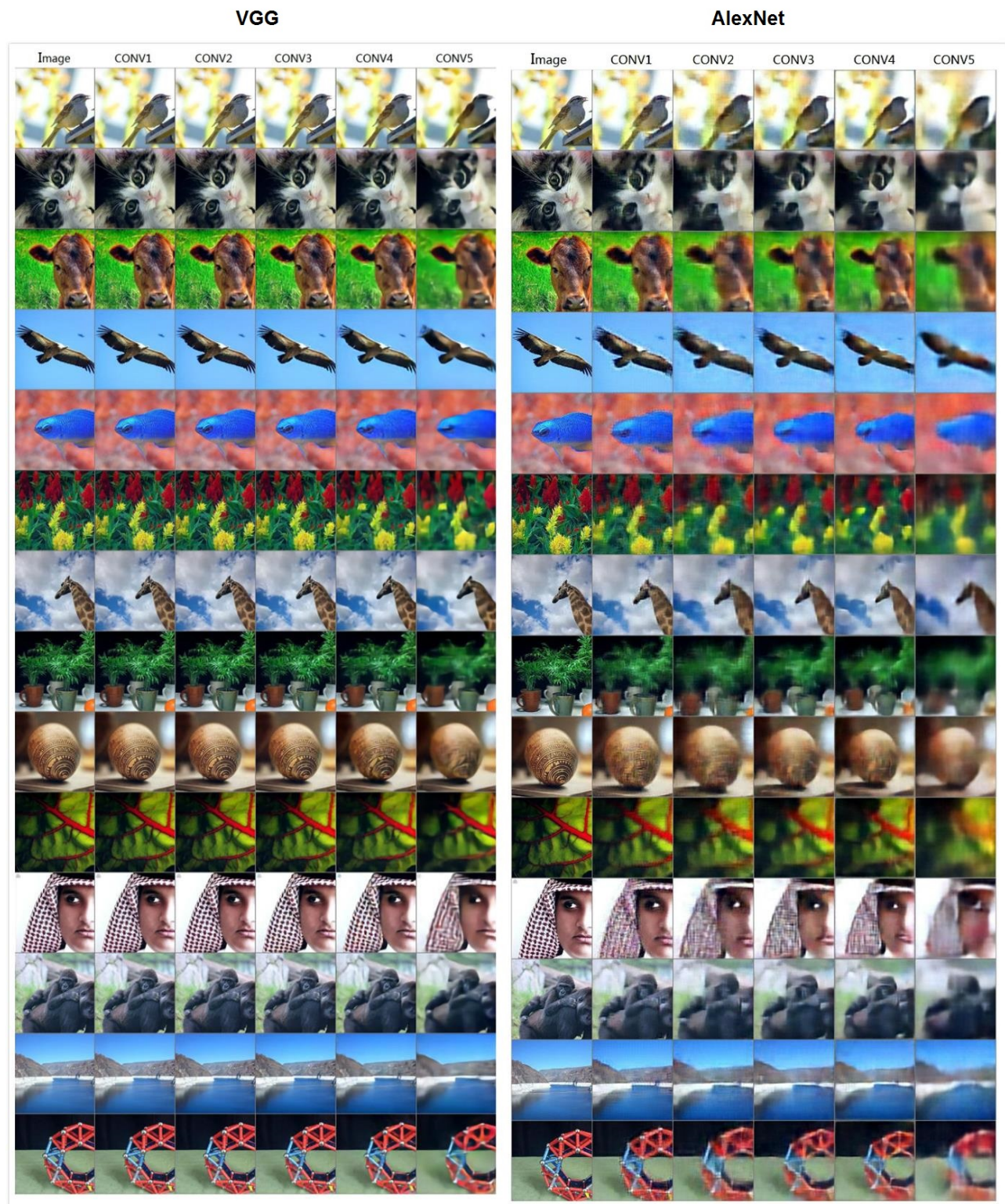


Figure 12: More image reconstruction on trained-DeCNN for random weight VGG16/AlexNet. Left column is for VGG16 and right column is for AlexNet. The reconstruction quality is higher on VGG16.